

La mobilità regionale degli studenti universitari

Opportunità e prime evidenze da una sperimentazione di linkage tra Anagrafe nazionale studenti e banca dati WHIP

Andrea Priulla
Università di Enna-Kore

Dalit Contini
Università di Torino

Sonia Della Monica
Università di Torino

Noemi Ferrara
Università di Torino

Il contributo presenta una sperimentazione di *linkage privacy-preserving* tra Anagrafe nazionale studenti e banca dati WHIP, finalizzata a integrare informazioni sui percorsi universitari e lavorativi. Il lavoro descrive la procedura di abbinamento, valuta le caratteristiche dell'archivio integrato risultante e ne illustra il potenziale informativo per lo studio della mobilità interregionale tra università e lavoro. I risultati preliminari evidenziano differenze territoriali nei percorsi dei laureati e degli studenti che interrompono gli studi.

Regional mobility of university students. Opportunities and preliminary evidence from a linkage procedure between the National student registry and the work histories Italian Panel
The paper presents a privacy-preserving linkage procedure between the National Student Registry and the WHIP database, aimed at integrating information on university and employment trajectories. It describes the matching procedure, assesses the characteristics of the resulting integrated archive, and illustrates its analytical potential for studying interregional mobility between university and work. The preliminary results highlight territorial differences in the trajectories of graduates and students who discontinue their studies.

DOI: 10.53223/Sinappsi_2026-02-7

Citazione	Parole chiave	Keywords
Priulla A., Contini D., Della Monica S., Ferrara N. (2026), La mobilità regionale degli studenti universitari. Opportunità e prime evidenze da una sperimentazione di linkage tra Anagrafe nazionale studenti e banca dati WHIP, <i>Sinappsi</i> , XVI, n.2, pp.106-121	Dati amministrativi Mobilità giovanile Studenti	Administrative data Youth mobility Students

Introduzione

La mobilità regionale degli studenti universitari e dei laureati è diventata negli ultimi anni uno dei temi centrali nel dibattito sulle disuguaglianze territoriali in Italia. Il trasferimento di giovani dalle regioni del Mezzogiorno verso i Poli metropolitani del Centro-Nord e verso l'estero per studio o per lavoro è spesso letto in chiave di 'fuga dei cervelli', con implicazioni importanti per le traiettorie di sviluppo regionale e per la progettazione delle politiche di istruzione e lavoro (Dotti *et al.* 2013; Priulla *et al.* 2025; Tosi *et al.* 2025).

Nonostante la rilevanza del tema, l'evidenza empirica disponibile sulla mobilità regionale dopo la laurea rimane ancora parziale. Attualmente, la principale fonte informativa è rappresentata dalle indagini AlmaLaurea, che combinano archivi amministrativi universitari con interviste rivolte all'intera popolazione dei laureati degli atenei che aderiscono al Consorzio, al momento della laurea e a 1, 3 e 5 anni dal conseguimento del titolo. Queste indagini offrono informazioni molto dettagliate sui corsi di studio, sulle aspirazioni e sulle prime esperienze di lavoro, e sono ampiamente utilizzate

per il monitoraggio e la valutazione delle politiche universitarie.

Tuttavia, i dati AlmaLaurea presentano alcuni limiti che rendono difficile l'analisi delle biografie di vita individuali. Anzitutto, l'orizzonte temporale delle informazioni è limitato a specifici momenti della vita degli individui, senza la possibilità di ricostruire le sequenze di episodi lavorativi, disoccupazione, inattività e gli eventuali spostamenti di regione nel medio periodo. In secondo luogo, si pone un problema di *attrition*, ovvero di riduzione del tasso di risposta alle indagini successive alla laurea, che diminuisce nel tempo e in modo non casuale: tendono infatti a uscire dal campione soprattutto i laureati triennali, introducendo potenziali distorsioni nelle stime di mobilità e degli esiti occupazionali. Infine, i dati AlmaLaurea consentono di osservare esclusivamente gli individui che hanno conseguito la laurea, mentre non raccolgono informazioni su chi non ha concluso il percorso universitario, perché ancora iscritto o perché ha abbandonato gli studi.

Accanto a questa fonte, l'Italia dispone di due grandi famiglie di archivi amministrativi potenzialmente cruciali per lo studio congiunto di istruzione terziaria e lavoro: l'Anagrafe nazionale degli studenti (ANS), che segue longitudinalmente le carriere di studio degli individui iscritti a tutte le università italiane e la banca dati WHIP (*Work and Health Italian Panel*), che ricostruisce – a partire dagli archivi Inps, Inail, del Ministero della Salute e del Ministero del Lavoro – le biografie di lavoro e salute di un campione rappresentativo di lavoratori in Italia, includendo in particolare le Comunicazioni Obbligatorie (COB) su attivazioni, trasformazioni e cessazioni dei contratti di lavoro, che contengono informazioni molto dettagliate sulla professione svolta.

Se i dati ANS potessero essere collegati a dati amministrativi di fonte Inps o Ministero del Lavoro (quali WHIP e COB), sarebbe possibile seguire gli individui dal luogo di nascita e di scuola superiore, al luogo di iscrizione universitaria, fino ai luoghi di lavoro effettivi, ricostruendo così percorsi di mobilità regionale in una prospettiva longitudinale. Tuttavia, gli archivi sono gestiti da soggetti diversi, sono stati costruiti per finalità amministrative distinte e attualmente non sono pensati per essere integrati a fini di ricerca. Pertanto, salvo alcuni esperimenti circoscritti a singoli atenei o regioni, mancano infrastrutture dati stabili che consentano di usare congiuntamente queste fonti su scala nazionale (Leombruni e Ardito 2026).

Questo articolo presenta una sperimentazione di integrazione tra l'archivio ANS e WHIP, con la finalità di esplorarne l'utilità in un'ottica di *policy analysis*. In particolare, ci si prefigge di: a) mostrare come,

con una procedura di *linkage privacy-preserving*, sia possibile mettere in relazione i due archivi amministrativi; b) discutere il potenziale di questa infrastruttura per lo studio della mobilità interna interregionale degli studenti e dei laureati nelle transizioni tra università e lavoro.

1. Quali archivi per quali potenzialità

La diffusione di dati amministrativi in ambito di ricerca economica e sociale è stata accompagnata da una crescita di studi che ne sfruttano la copertura di popolazione, la profondità temporale e la relativa economicità rispetto alle indagini campionarie. In Italia, diversi lavori impiegano i dati derivanti dall'Anagrafe nazionale studenti (ANS), per studiare la mobilità di studio e le carriere degli studenti universitari (D'Agostino *et al.* 2019; Attanasio *et al.* 2020; Usala *et al.* 2023; Priulla *et al.* 2025); altri lavori utilizzano i dati di fonte Inps per approfondire molteplici aspetti del mercato del lavoro (Berton *et al.* 2012; Galizzi *et al.* 2023; Devicienti *et al.* 2025; Ardito *et al.* 2026 solo per fare alcuni esempi). Al tempo stesso, la riflessione metodologica su come questi dati vengono generati, integrati e interpretati è ancora meno strutturata rispetto a quella maturata attorno alle indagini statistiche tradizionali (Leombruni 2024).

Ad oggi, infatti, la principale fonte dati utilizzata per esplorare il percorso universitario e lavorativo dei laureati italiani è l'indagine AlmaLaurea, che offre una base informativa sulle caratteristiche dei corsi, delle carriere universitarie e dei primi esiti occupazionali. Tuttavia, come discusso nella sezione precedente, oltre a riguardare esclusivamente coloro che conseguono il titolo universitario, i dati AlmaLaurea presentano diversi limiti che rendono difficile affrontare domande di ricerca in una prospettiva di biografia di vita, ovvero seguire nel tempo le traiettorie di studio e lavoro degli individui.

In particolare, l'indagine raccoglie informazioni (autoriportate) in specifici momenti nel tempo – al momento della laurea e successivamente a 1, 3 e 5 anni – senza permettere una ricostruzione continua dei percorsi individuali. A questo si aggiunge un problema di *attrition* non casuale che riduce, e presumibilmente distorce, l'informazione disponibile sulle traiettorie degli individui: le indagini successive a quella del momento della laurea presentano infatti tassi di risposta progressivamente più bassi, con una caduta particolarmente marcata tra i laureati triennali (Appendice, tabella A1). Quest'ultima è legata principalmente al disegno dell'indagine: nel questionario somministrato a 1, 3 e 5 anni dalla laurea triennale viene chiesto se il rispondente si sia iscritto a un corso magistrale. In caso di risposta affermativa, il questionario si interrompe e l'intervistato viene contattato nuovamente solo al

conseguimento del titolo magistrale. Non è quindi possibile osservare eventuali episodi di lavoro che si verificano durante il percorso di laurea magistrale e se lo studente abbandona gli studi nel corso della magistrale, viene perso dall'osservazione.

In aggiunta alla difficoltà di ricostruire le biografie di lavoro, un ulteriore limite dei dati AlmaLaurea è rappresentato dalla natura delle informazioni sulla professione, che nelle interviste a 1 e 3 anni seguono una classificazione ad hoc e con un grado di dettaglio limitato, mentre solamente nell'intervista a 5 anni si fa riferimento alla codifica Istat a 5 digit.

In questo contesto, gli archivi amministrativi longitudinali di fonte Inps e Ministero del Lavoro potrebbero rappresentare una risorsa preziosa per seguire intere coorti di individui nel tempo, osservando il percorso di studio – tramite le informazioni su regione di frequenza, tipologia del diploma, percorso accademico inclusi eventuali cambi di corso di laurea e abbandoni – e i vari episodi di lavoro, con indicazioni puntuali sulla professione svolta, nonché eventuali periodi di disoccupazione e episodi di mobilità territoriale.

Il problema, tuttavia, è che attualmente gli archivi amministrativi relativi a università e lavoro non sono integrati in modo sistematico. L'ANS fornisce informazioni dettagliate sui percorsi accademici, ma non dice nulla sulla carriera occupazionale. Allo stesso tempo WHIP e COB descrivono le carriere lavorative, ma includono la sola informazione sul massimo titolo di studio conseguito, che è tuttavia incompleta e di qualità molto incerta; inoltre, non vengono registrati né il luogo di conseguimento del titolo né il corso di laurea o l'area disciplinare.

In questo contesto, la sperimentazione si propone come un caso concreto di integrazione di fonti amministrative, basato su una procedura di *linkage privacy-preserving* che consente di mettere in relazione archivi di natura diversa. L'obiettivo non è soltanto sfruttare dati già raccolti, ma mostrare come la loro integrazione possa consentire di rispondere a nuovi interrogativi di ricerca. In particolare, esploriamo qui su un tema – la mobilità interregionale degli studenti (laureati e non) tra università e lavoro – che potrebbe beneficiare sostanzialmente dall'analisi di dati amministrativi longitudinali.

L'Anagrafe nazionale degli studenti

Le fonti di dati impiegate dal nostro gruppo di ricerca per l'esecuzione di questa sperimentazione

sono l'Anagrafe nazionale degli studenti e WHIP. L'ANS è l'archivio amministrativo che registra le carriere di tutti gli studenti iscritti alle università italiane, fornendo informazioni anche sul percorso scolastico pregresso, quali la tipologia di diploma, nonché l'anno di conseguimento e l'area geografica dell'istituto frequentato. Va tuttavia segnalata una limitazione dei dati dell'ANS, rappresentata dall'assenza di informazioni sugli individui che si iscrivono in università all'estero, o che si trasferiscono all'estero dopo l'immatricolazione nel sistema universitario italiano.

Per ciascun individuo sono disponibili i seguenti dati:

- caratteristiche anagrafiche (sesso, data, Comune e regione di nascita, cittadinanza ecc.);
- informazioni sulla scuola superiore (tipo di diploma, anno di conseguimento, area geografica dell'istituto, voto di diploma ecc.);
- dettagli a cadenza annuale sulle carriere universitarie (università, corso di studio, tipo di laurea, data di immatricolazione, crediti formativi universitari (CFU), eventuali passaggi di corso o ateneo, data di conseguimento della laurea, interruzioni).

I dati dell'ANS sono stati resi disponibili grazie alla stipula di un protocollo di ricerca tra il MUR, Invalsi e alcune università con capofila l'Università di Palermo, nell'ambito di un progetto PRIN finalizzato allo studio della mobilità di studio dei giovani italiani¹. L'utilizzo di questi dati (confluiti in un archivio denominato Mobysu.it, prodotto nel 2017 e in seguito aggiornato periodicamente) ha consentito di superare le limitazioni degli archivi dei singoli atenei, che non permettono di seguire gli studenti in caso di trasferimento tra università.

Per la sperimentazione di *linkage* sono stati impiegati i dati relativi alle coorti di immatricolati in tutte le università italiane negli anni accademici dal 2009/10 al 2016/17 (con l'esclusione dell'anno 2011 dovuta a problemi di dati mancanti sul Comune di nascita degli studenti). Ciascuna coorte viene osservata fino al 30 giugno 2019. Nello specifico, la sottopopolazione dei dati dell'ANS utilizzata è costituita da 96.968 individui, derivata dall'applicazione dello stesso schema di campionamento impiegato per WHIP, ovvero uno schema di campionamento sistematico basato su 2 date di nascita per ogni mese, a seguito della quale

1 I dati utilizzati in questo studio sono stati elaborati in conformità al protocollo di ricerca per lo studio *From high school to the job placement: analysis of university careers and university mobility from Southern to Northern Italy* tra il Ministero dell'Università e della Ricerca, il Ministero dell'Istruzione e del merito, l'Università degli Studi di Palermo come ente capofila, e l'Istituto Invalsi. Sono coinvolte inoltre le università di Cagliari, Firenze, Napoli Federico II, Sassari, Siena, Torino (in seguito anche Enna KORE, Salerno, Bologna).

sono state effettuate diverse operazioni di selezione. In primo luogo, sono stati considerati coloro la cui età al momento della prima immatricolazione risultava non superiore a 22 anni. Sono stati esclusi anche tutti gli studenti che, in un qualsiasi momento della propria carriera universitaria siano risultati iscritti ad un'università telematica. Questa scelta risiede nel fatto che per frequentare un corso di laurea offerto da un'università telematica non è necessario trasferirsi nella regione sede dell'ateneo. Infine, è stato rimosso un numero limitato di record che presentavano anomalie strutturali nelle carriere (duplicazioni di ID, sequenze di immatricolazioni e lauree logicamente impossibili, sovrapposizioni non risolubili tra corsi diversi), per cui non era possibile ricostruire in modo credibile il percorso di studio. Per concludere, sono stati esclusi gli studenti immatricolati ad un corso di laurea a ciclo unico, poiché costituiscono percorsi al cui termine si aspira generalmente ad un percorso lavorativo che richiede l'iscrizione ad un albo o ad una professione normata, non incluse all'interno del database WHIP.

WHIP e l'archivio delle Comunicazioni Obbligatorie

Il secondo archivio di dati amministrativi utilizzato per la sperimentazione di *linkage* è WHIP, una banca dati che ricostruisce le biografie lavorative e di salute di un campione di circa il 7% dei lavoratori italiani (tramite l'estrazione su due date mensili di nascita), il cui titolare è il Ministero della Salute ed è costruita a partire da diversi archivi amministrativi². Le fonti che compongono questo database sono: l'Inps con dati sugli episodi lavorativi relativi alla maggior parte delle forme di lavoro in Italia (dipendenti, parasubordinati, professionisti, artigiani e commercianti), nonché le dichiarazioni ISEE e le informazioni su benefici di welfare (quali pensioni o indennità di disoccupazione) e il Ministero del Lavoro con le Comunicazioni Obbligatorie (COB), che forniscono informazioni dettagliate sui contratti di lavoro subordinato e parasubordinato pubblico e privato dal 2008 al 2019 (ultimo aggiornamento a nostra disposizione)³.

Complessivamente, nei dati WHIP sono comprese:

- tutte le fattispecie di lavoro registrate nelle COB, incluse collaborazioni non occasionali, specializzandi e dottorandi;

- Partite IVA (artigiani, professionisti, commercianti non con cassa separata).

Non sono compresi invece albi a gestione separata e alcune forme di lavoro occasionale.

Date le finalità dello studio, tra tutte le informazioni presenti nella banca dati WHIP sono state considerate esclusivamente le variabili anagrafiche di base (sesso, data di nascita, Comune di nascita) e le principali informazioni sui rapporti di lavoro (inizio e fine contratto, regione di lavoro, classificazione della professione). Rispetto al campione di individui selezionati per il *linkage*, sono stati considerati solo individui appartenenti alle stesse coorti selezionate dal lato ANS, nati in Italia e che abbiano presentato almeno una dichiarazione ISEE. Le ultime due selezioni sono motivate dal fatto che nella chiave di abbinamento è incluso il Comune di nascita, informazione disponibile solo per gli individui nati in Italia ed esclusivamente sul flusso informativo ISEE⁴.

2. Il disegno di linkage

Nel seguito utilizziamo il termine *linkage* per indicare la procedura complessiva di integrazione tra il dataset Mobysu.it di fonte ANS e il dataset WHIP. Con il termine *match* indichiamo invece l'esito della procedura, ossia una possibile corrispondenza tra uno o più record dei due archivi. Come sarà chiarito nei paragrafi successivi, la procedura adottata in questa fase si basa su una chiave di abbinamento che *non* consente l'identificazione certa dei soggetti (sesso, Comune e data di nascita).

Entrambi gli archivi, ancorché privi di identificativi diretti, contengono variabili che – soprattutto se combinate tra loro – possono aumentare il rischio di reidentificazione: può infatti accadere che un individuo presenti una combinazione unica di caratteristiche che ne consente una potenziale identificazione indiretta. Nonostante in questo studio si consideri una chiave di abbinamento basata su un insieme minimale di informazioni, è stato comunque adottato, per le procedure di *linkage*, un disegno a doppio cieco, costituito da due gruppi di lavoro nettamente distinti per compiti e informazioni accessibili.

La distinzione tra gruppi di lavoro risponde a un principio di minimizzazione delle informazioni

2 Vedi <https://www.epi.piemonte.it/whipsalute/>.

3 WHIP include anche dati sugli infortuni sul lavoro e le malattie professionali (Inail), e le Schede di dimissione ospedaliera rilasciate dal Ministero della Salute.

4 Ha presentato una dichiarazione ISEE il 46% dei lavoratori delle coorti di nascita di interesse presenti nel file WHIP. Tuttavia, è probabile che tra gli studenti universitari questa quota sia sensibilmente più elevata, date le forti riduzioni nelle tasse universitarie previste per i redditi medi e bassi. Nel complesso, è plausibile che il campione analizzato sia maggiormente rappresentativo di una popolazione con livelli di reddito non elevati.

accessibili a ciascuno: chi effettua il *linkage* iniziale non dispone delle informazioni sostantive sulle carriere, mentre chi analizza le biografie non accede alle variabili anagrafiche disaggregate utilizzate per costruire la chiave di *linkage*.

Il Team A ha operato su due file, uno derivante dal database ANS e l'altro da WHIP, senza però sapere quale dei due provenisse da quale archivio, ognuno composto esclusivamente da:

- un identificativo interno anonimo;
- sesso;
- data di nascita completa (giorno, mese, anno);
- Comune di nascita.

Su questi due file il Team A ha applicato una regola di *linkage* basata sull'esatta coincidenza di sesso, data di nascita e Comune di nascita, identificando tutti i possibili gruppi di record che condividevano la stessa chiave. Il dataset risultante è quindi composto dalle diverse possibili configurazioni. Match univoci (match 1:1): un record ANS e un record WHIP; match non univoci: un record ANS associato a più record WHIP (1:molti); più record ANS associati a un unico record WHIP (molti:1); più record ANS collegati a più record WHIP (molti:molti). Inoltre vi sono record ANS che non vengono abbinati ad alcun record WHIP.

Poiché Comune e data di nascita sono le informazioni che contribuiscono maggiormente al rischio di reidentificazione, ma non sono necessarie all'analisi, il Team A restituisce al Team B un file in cui tali informazioni sono riportate a un livello maggiore di aggregazione. In particolare, il file contiene: identificativi anonimi, sesso, regione e anno di nascita.

Su questo file, il Team B associa, per ciascun identificativo, un set di variabili longitudinali:

- per i record provenienti da ANS: informazioni sintetiche su scuola superiore, università, tipo di corso, data di immatricolazione e di laurea, eventuali cambi di corso/ateneo;
- per i record provenienti da WHIP: informazioni sui rapporti di lavoro (periodo di inizio/fine, tipologia contrattuale, regioni di lavoro, descrizione della professione).

In questo modo, il Team B dispone di un insieme di biografie di studio e lavoro organizzate in diversi gruppi di match, senza tuttavia disporre delle variabili chiave nella loro forma disaggregata. Ciò consente di tutelare la privacy e, al tempo stesso, di

valutare il potenziale informativo dei dati.

Entriamo ora nel merito dei dati così prodotti.

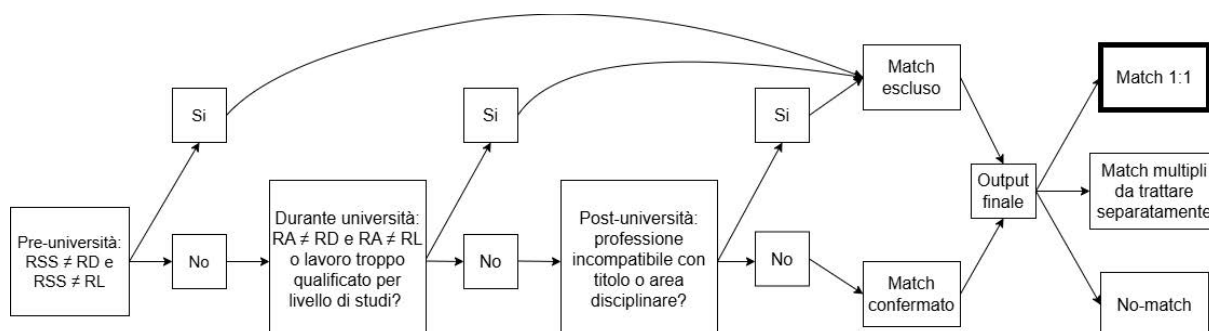
- 1) In primo luogo, come si può interpretare l'assenza di un match valido? Lo scenario più plausibile è che l'individuo non abbia avuto esperienze lavorative. Tuttavia, non si possono escludere altre casistiche: ad esempio, l'individuo potrebbe lavorare all'estero, svolgere un'attività non ricompresa tra le fattispecie registrate in WHIP, oppure non avere mai presentato una dichiarazione ISEE (in questo caso, come descritto sopra, l'informazione sul Comune di nascita, che costituisce parte della chiave di *linkage*, non è disponibile dal lato WHIP).
- 2) Nel caso di match 1:1, l'abbinamento tra un solo studente – identificato da una combinazione univoca di sesso, data di nascita e Comune di nascita – e un solo lavoratore non garantisce che la biografia individuale sia ricostruita con certezza. Nello stesso giorno e luogo potrebbe infatti essere nato un individuo con esperienza lavorativa ma che non ha frequentato l'università (e quindi non risulta nell'archivio ANS), mentre lo studente potrebbe aver lavorato all'estero o non essere ancora entrato nel mercato del lavoro (e quindi non compare in WHIP). Ciò comporta che, sebbene la procedura di *linkage* sia basata su regole esatte di corrispondenza tra chiavi di aggancio predefinite, nessun match sia configurabile come risultato di un abbinamento certo, perché più persone possono condividere la stessa combinazione di sesso, data e Comune di nascita. La probabilità che i due record abbinati univocamente si riferiscano alla stessa persona decresce all'aumentare della dimensione del Comune, perché diventa via via più plausibile che vi siano più persone nate nella stessa data⁵.
- 3) Anche nei match 1:molti o molti:1 – che, come evidenziato nelle conclusioni, costituiranno l'oggetto di un proseguimento di questo lavoro tramite abbinamento probabilistico – non è certo che tra i record presenti nel lato 'molti' vi sia effettivamente quello riferito alla stessa persona presente nel lato singolo.

3. Regole di coerenza biografica per l'affinamento dei match

L'affidabilità dei match può essere ulteriormente migliorata mediante l'applicazione di alcune regole di esclusione e di coerenza biografica, organizzate

5 Condizionatamente all'osservazione di un match 1:1, per i laureati la probabilità che i due record si riferiscano allo stesso individuo è pari a circa il 98% in un Comune di 10.000 abitanti e al 70% in un Comune di 100.000 abitanti. Tale stima è stata ottenuta assumendo un tasso di natalità pari a quello medio italiano del periodo, insieme a ipotesi semplificatrici sulla probabilità di conseguire una laurea e di essere occupati tra i laureati. I risultati variano solo marginalmente al variare di tali ipotesi.

Figura 1. Flow-chart delle regole di affinamento applicate alle coppie di match studente-lavoratore generati dalla procedura di linkage



RSS = Regione scuola superiore; RD = Regione domicilio; RL = Regione lavoro; RA = Regione ateneo.

Fonte: elaborazione degli Autori

lungo l'asse temporale delle carriere di studio (prima dell'università, durante l'università, dopo l'università) e integrate con informazioni ausiliarie sui percorsi lavorativi, provenienti dalle COB. Le regole di esclusione hanno lo scopo di identificare coppie di record associate con informazioni incoerenti tra loro, articolate in tre fasi temporali: 1) precedente l'immatricolazione ad un corso di laurea; 2) durante il periodo universitario; 3) successiva al conseguimento di un titolo di laurea, triennale o magistrale.

Uno schema sintetico dei passaggi effettuati è visibile in figura 1.

Nella fase pre-universitaria utilizziamo l'informazione, presente in ANS, relativa alla regione in cui è stato conseguito il diploma di scuola superiore (RSS) e la confrontiamo con le informazioni provenienti da WHIP sulla regione di domicilio (RD) e sulla regione in cui si sono svolti gli eventuali episodi di lavoro precoci (RL). Un match viene escluso quando la regione della scuola superiore è diversa sia dalla regione di domicilio sia dalla regione di lavoro, ossia nei casi in cui valga contemporaneamente $RSS \neq RD$ e $RSS \neq RL$. La logica sottostante è la seguente: dato uno studente x e un lavoratore y che condividono la stessa chiave anagrafica di *linkage*, è più probabile che siano la stessa persona se, nel periodo pre-universitario, la regione in cui x frequenta le superiori coincide con la regione in cui y vive e lavora, dal momento che l'eventuale lavoro viene svolto in contemporanea con gli studi. Quando questa condizione di coerenza territoriale non è soddisfatta, la coppia viene scartata perché ritenuta poco plausibile.

Nella fase relativa agli anni di iscrizione ad un corso di laurea, applichiamo una regola analoga, sostituendo alla regione della scuola superiore la

regione dell'ateneo (RA). Vengono esclusi i match in cui la regione dell'università è diversa sia dalla regione di domicilio sia dalla regione di lavoro, cioè i casi in cui $RA \neq RD$ e $RA \neq RL$. La logica è molto simile a quella della regola precedente: se studio, residenza e (eventuale) lavoro si collocano per l'intero periodo in regioni completamente diverse, senza alcuna sovrapposizione territoriale, il match è ritenuto non plausibile.

A questa regola generale si aggiunge un controllo specifico sulla coerenza fra livello di studi e tipo di occupazione: vengono esclusi gli studenti che hanno abbandonato il percorso universitario e che risultano aver svolto un lavoro per il quale è richiesta una laurea, senza aver conseguito almeno un titolo triennale (con eccezione degli studenti vicini alla laurea, che potrebbero essere stati assunti come laureati perché in procinto di ottenerla). Nel complesso, l'applicazione delle regole per la fase universitaria e pre-universitaria seleziona un sottoinsieme di match 1:1 che mantiene la coerenza territoriale fra studi, domicilio e lavoro e la compatibilità fra livello di studi effettivamente raggiunto e qualificazione del lavoro svolto.

La fase post-universitaria si concentra sulla coerenza tra titolo di studio e occupazione dopo l'uscita dall'università. Le esclusioni si basano su criteri più puntuali di congruenza tra laurea e tipo di lavoro, con particolare attenzione alle professioni regolamentate o per le quali è richiesta una specifica formazione universitaria. L'idea di fondo è che non sia plausibile che un individuo presente in ANS senza il titolo richiesto coincida con un lavoratore WHIP occupato in una professione per cui quel titolo è obbligatorio: ad esempio, chi lavora come veterinario deve avere la laurea in medicina veterinaria; in caso contrario, è verosimile che si tratti di due persone diverse.

Operativamente, viene applicata una serie di filtri. Si eliminano innanzitutto i casi in cui compaiono lavori in area sanitaria senza una laurea medica. Vengono poi esclusi i docenti privi del titolo adeguato e i laureati triennali che risultano docenti universitari. Un controllo specifico riguarda gli studenti iscritti a corsi triennali che risultano contemporaneamente ricercatori o docenti, così come i non laureati occupati in posizioni accademiche (ricercatori, docenti o altre figure universitarie).

Ulteriori filtri riguardano la coerenza tra area disciplinare e professioni ad alta specializzazione tecnica: vengono ad esempio esclusi i laureati triennali in ambito umanistico che risultano lavorare come ingegneri per periodi superiori a 30 giorni. Un ulteriore filtro elimina i casi di dottorato senza laurea: in queste situazioni, qualora la laurea fosse stata conseguita all'estero, l'individuo non sarebbe comunque osservabile in ANS.

Infine, si applicano controlli sulle professioni per le quali è richiesta l'iscrizione ad albi (ingegneri, commercialisti e altre professioni ordinistiche), escludendo le coppie in cui il lavoratore svolge un'attività soggetta ad albo ma non possiede in ANS una laurea coerente.

A seguito di queste operazioni, l'insieme iniziale di 96.968 studenti si suddivide come segue: 16.777 persone in match univoci 1:1 (17,3%), 10.117 in match 1:molti (10,4%), 16.252 in match molti:1 (15,9%), 15.453 in match molti:molti (15,9%) e 38.369 studenti senza match (39,6%).

4. Caratteristiche del dataset 1:1

I gruppi di studenti in match non univoci saranno oggetto di analisi approfondita in un prossimo futuro, per verificare se e come sia possibile assegnare a ciascuno studente una o più biografie lavorative plausibili mediante l'adozione di un approccio probabilistico. In questo lavoro ci si concentra invece sulla adeguatezza dell'utilizzo del dataset dei match 1:1 per specifiche finalità di ricerca.

Come evidenziato sopra, in questo archivio la distribuzione degli individui per dimensione del Comune di nascita è sbilanciata verso i nati in Comuni di piccola dimensione. In tabella 1 è mostrata la distribuzione degli studenti che risultano nel gruppo 1:1, nel gruppo di match non univoci e tra i no-match, per dimensione del Comune. In linea con le aspettative, gli studenti nati in Comuni di piccole dimensioni risultano sovrarappresentati tra i 1:1 e i no match rispetto a quelli nati nei grandi centri, poiché in questi Comuni la chiave utilizzata per il *linkage* è più discriminante. Più di un terzo dei match 1:1 riguarda infatti individui nati in Comuni fino ai 25.000 abitanti, mentre tra i match non univoci questa percentuale è pari al 7% circa.

Si effettua quindi un'analisi esplorativa per valutare la rappresentatività del campione 1:1 rispetto alla popolazione di riferimento, sulla base di alcune caratteristiche tipicamente associate alla carriera lavorativa. Idealmente, si vorrebbero confrontare gli studenti nel gruppo 1:1 con gli altri studenti con esperienze di lavoro. L'identificazione

Tabella 1. Distribuzione degli studenti per dimensione del Comune di nascita nei diversi gruppi di match

Dimensione del Comune	Match univoci (1:1)		Match non univoci		No match	
	N	%	N	%	N	%
< 10 mila	1.815	10,8	699	1,7	3.950	10,3
10 – 25 mila	3.875	23,1	2.247	5,4	8.523	22,2
25 – 50 mila	4.906	29,2	4.298	10,3	10.448	27,2
50 – 100 mila	3.374	20,1	5.511	13,2	7.482	19,5
100 – 250 mila	1.664	9,9	6.222	14,9	4.410	11,5
250 – 500 mila	290	1,7	3.906	9,3	1.074	2,8
500 mila – 3 milioni	72	0,4	17.214	41,2	889	2,3
NA	781	4,7	1.725	4,1	1.593	4,2
TOT	16.777	100,0	41.822	100,0	38.369	100,0

Fonte: elaborazione degli Autori

Tabella 2. Distribuzione del percorso universitario nei diversi gruppi di match

Stato del percorso	Match univoci (1:1)		Match non univoci	
	N	%	N	%
Laurea Triennale in corso	4.416	26,3	11.653	27,9
Laureato Triennale	3.178	18,9	8.293	19,8
Laurea Magistrale in corso	2.319	13,8	5.609	13,4
Laureato Magistrale	2.639	15,7	5.828	13,9
Abbandono di Laurea Magistrale	234	1,4	624	1,5
Abbandono di Laurea Triennale	3.879	23,1	9.489	22,7
Passaggio a corso a Ciclo Unico	112	0,7	326	0,8

Fonte: elaborazione degli Autori

Tabella 3. Distribuzione dell'area disciplinare nei diversi gruppi di match

Area disciplinare di laurea	Match univoci (1:1)		Match non univoci	
	N	%	N	%
Benessere e Salute	1.407	9,2	4.001	10,6
Economico-Giuridica	1.599	10,4	4.213	11,1
Formazione	429	2,8	867	2,3
Ingegneria e Architettura	1.556	10,1	3.568	9,4
Politico-Sociale	757	4,9	2.064	5,5
Scienze e Tecnologia	1.371	8,9	2.784	7,4
Umanistico-Linguistica	1.448	9,4	3.346	8,8
Non laureati	8.210	53,4	20.979	55,5

Fonte: elaborazione degli Autori

certa di quest'ultimo gruppo non è tuttavia possibile, perché la procedura di abbinamento effettuata non garantisce abbinamenti esatti. Come prima approssimazione, confrontiamo il gruppo 1:1 con l'insieme degli studenti che presentano almeno un match. Escludiamo quindi i no-match, che è plausibile che non abbiano avuto esperienze lavorative.

Le tabelle 2 e 3 mostrano le distribuzioni relative allo stato del percorso universitario (laurea triennale o magistrale, studente in corso, abbandono del percorso di studi) e all'area disciplinare del corso di laurea. Si rileva una lieve sovrarappresentazione di laureati magistrali nel gruppo 1:1, e di alcune aree disciplinari come Ingegneria e Scienze e Tecnologia.

Per analizzare se il sottoinsieme dei match 1:1 differisca sistematicamente dal resto degli individui abbinati con un approccio multivariato,

sono stati stimati modelli di regressione logistica sulla probabilità di appartenere al gruppo 1:1. La variabile dipendente è un indicatore binario che assume valore 1 se l'individuo (lato ANS) presenta un match 1:1 e valore 0 se ricade in un gruppo di abbinamento non univoco. L'analisi viene dunque condotta sugli studenti che presentano almeno un match nel lato WHIP. Come variabili esplicative sono state considerate: la coorte di immatricolazione, il genere, il tipo di diploma di scuola media superiore, l'area disciplinare, lo stato del percorso universitario e la dimensione del Comune.

I risultati, riportati in appendice (tabella A2), mostrano che – con l'eccezione della dimensione del Comune di nascita⁶ – la maggior parte delle variabili non presenta effetti statisticamente significativi sulla probabilità di appartenere al gruppo 1:1, oppure presenta effetti significativi ma di entità molto contenuta. Inoltre, si osserva una

6 Essere nati in Comuni più grandi riduce la probabilità di appartenere al gruppo 1:1 di diverse decine di punti percentuali rispetto alla categoria di riferimento (meno di 10.000 abitanti).

modesta differenza di genere: gli studenti maschi hanno una probabilità leggermente più elevata di appartenere al gruppo 1:1 (circa 3 punti percentuali in più rispetto alle studentesse), in linea con una maggiore presenza maschile in alcune professioni regolamentate, incluse nelle regole di esclusione riguardanti la fase post-universitaria. *Ceteris paribus*, si osserva una probabilità di appartenenza al gruppo 1:1 lievemente più elevata per le aree disciplinari Formazione e Scienze e tecnologie, mentre non si confermano le differenze relative al diverso percorso universitario.

La conclusione preliminare che possiamo trarre è che il *linkage* 1:1 non individui un gruppo di studenti molto diverso dalla popolazione di partenza, nonostante le selezioni introdotte dalla procedura utilizzata e dalla struttura degli archivi amministrativi. Ulteriori analisi dovranno essere condotte per verificare la rappresentatività del campione in termini di storie di lavoro.

È opportuno, tuttavia, ricordare che alcuni meccanismi di selezione sono strutturali al disegno del *linkage* e alla natura stessa degli archivi utilizzati. In particolare, il *linkage* privilegia implicitamente le persone che hanno avuto rapporti di lavoro regolari in settori coperti da WHIP e, dunque, effettivamente visibili nelle Comunicazioni Obbligatorie. Inoltre, risultano abbinati solo gli individui che hanno presentato almeno una dichiarazione ISEE nel periodo analizzato, perché nel lato WHIP le informazioni su data e comune di nascita sono presenti solo per questi. Tali elementi non sembrano compromettere la somiglianza complessiva tra popolazione generale e gruppo dei match 1:1, ma va ricordato che il campione ottenuto non è pienamente rappresentativo. Di conseguenza, questi meccanismi di selezione devono essere tenuti esplicitamente in conto nelle analisi empiriche attraverso analisi di sensitività e un'interpretazione dei risultati sempre ancorata ai vincoli di copertura propri dei dati amministrativi utilizzati.

5. Potenziale informativo per lo studio della mobilità interna tra università e lavoro

In questa sezione si intende fornire un'analisi preliminare della mobilità territoriale degli studenti universitari e dei laureati italiani. I risultati presentati hanno carattere illustrativo e sono finalizzati a illustrare le potenzialità del dataset costruito con la procedura di *linkage* descritta nelle sezioni precedenti.

Nel complesso, la combinazione di analisi descrittive e modelli suggerisce che il sottoinsieme dei match 1:1, pur segnato da selezioni strutturali legate alla dimensione del Comune di nascita, non presenta

forti distorsioni rispetto alle principali caratteristiche degli studenti. Ciò permette di utilizzare il dataset 1:1 come base per analisi sulla mobilità regionale e sulle transizioni studio-lavoro, a condizione di tenere esplicitamente conto, nelle interpretazioni e nelle eventuali ponderazioni, dei meccanismi di selezione legati alla geografia demografica e alla copertura degli archivi amministrativi.

Grazie alla disponibilità di informazioni sulla regione di nascita, regione della scuola superiore, regione dell'ateneo e tipo/area di laurea, nonché delle sequenze di episodi lavorativi (con regioni di lavoro, settori, gruppi professionali), diventa possibile esplorare in modo più approfondito un tema centrale quale la mobilità degli studenti e dei laureati italiani, prima e dopo il conseguimento di un titolo, e i relativi percorsi professionali, offrendo una lettura più ricca dei processi di selezione territoriale rispetto a quanto consentito dalle sole indagini campionarie.

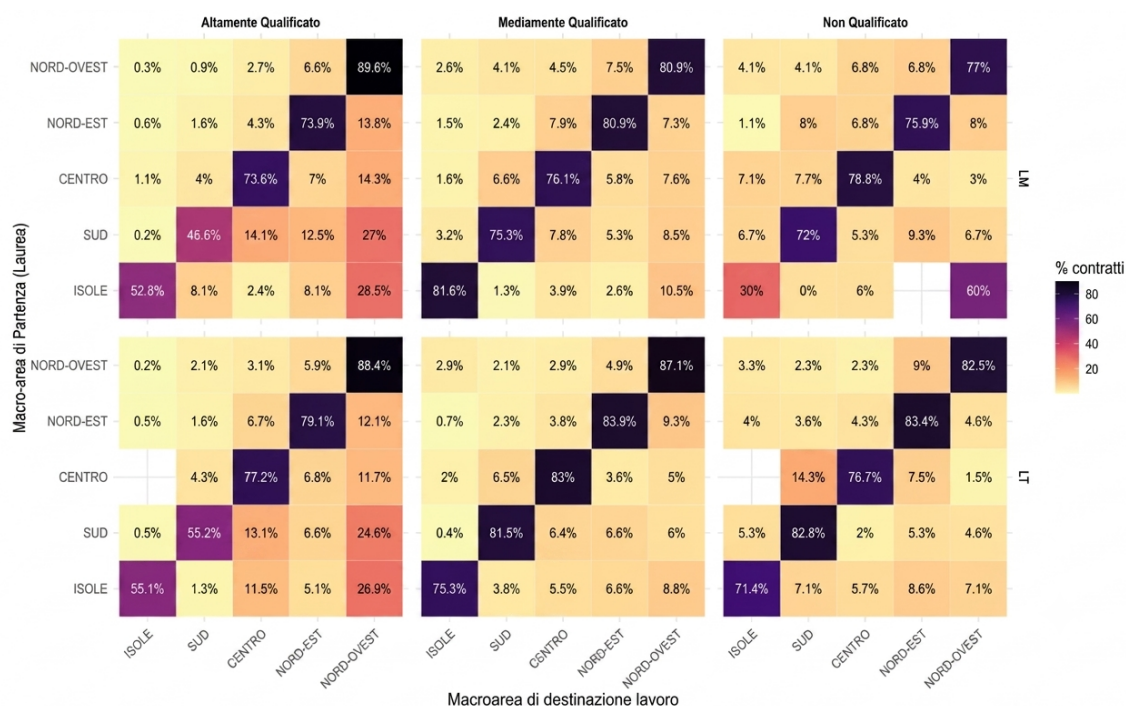
Inoltre, poiché nel dataset ANS sono presenti sia laureati sia studenti che hanno abbandonato il percorso universitario, è possibile confrontare le loro traiettorie, interrogandosi, ad esempio, se i non laureati entrino più precocemente nel mercato del lavoro ma con percorsi più localizzati e meno mobili; o se i laureati risultino più mobili territorialmente nella ricerca di un'occupazione; o ancora se vi siano differenze di mobilità e di qualità dell'occupazione tra laureati e non laureati. Si tratta di questioni centrali per la valutazione delle politiche legate al diritto allo studio e per l'analisi delle disuguaglianze economiche e sociali a livello territoriale.

Risultati preliminari

Il focus dell'analisi è la mobilità tra università e lavoro tra macro-aree. Tra i laureati magistrali, il 30% presenta almeno uno spostamento post-laurea rispetto alla macro-area dell'ateneo, mentre il 70% rimane stabilmente nella macro-area dell'ateneo. La mobilità è più contenuta tra i laureati triennali: il 23% sperimenta almeno uno spostamento, mentre il 77% rimane nell'area dell'ateneo. Il differenziale tra laureati magistrali e laureati triennali suggerisce che livelli di istruzione più elevati siano associati a una maggiore mobilità territoriale. La laurea magistrale sembra dunque ampliare l'orizzonte geografico del mercato del lavoro di riferimento.

Per analizzare più in dettaglio i singoli spostamenti, si mostrano di seguito le matrici di transizione dalla macro-area di conseguimento del titolo di studio universitario per i laureati (figura 2) o dell'ultimo ateneo frequentato per coloro che hanno abbandonato gli studi (figura 3) alla macro-area in cui ha sede l'azienda. Le unità statistiche sono qui costituite dai rapporti di lavoro;

Figura 2. Quote percentuali dei rapporti di lavoro per macro-area di partenza (sede dell'ateneo) e di arrivo (luogo di lavoro), per tipologia di laurea conseguita e livello di qualificazione del lavoro (classificazione ISCO)



Nota: Panel superiore: laurea magistrale, panel inferiore: laurea di primo livello. Coorti di immatricolazione 2009/10-2016/17. Rapporti di lavoro osservati negli archivi WHIP/COB fino al 2019.

Fonte: elaborazione degli Autori

pertanto, ogni individuo può essere coinvolto in più transizioni⁷. Le matrici sono articolate per livello di qualifica dell'occupazione, sulla base del primo digit del codice ISCO: occupazioni altamente qualificate (codici 1-3), mediamente qualificate (codici 4-6) e non qualificate (codici 7-9).

L'analisi degli spostamenti rivela che la propensione alla mobilità non è neutra rispetto alla qualità dell'inserimento occupazionale: gli spostamenti sono più frequenti nei segmenti professionali medio-alti, coerentemente con l'ipotesi che la mobilità sia parte di strategie di investimento in capitale umano e accesso a mercati del lavoro più dinamici.

Emergono inoltre forti differenze tra le diverse aree geografiche, in particolare per i laureati. Mentre le regioni del Centro e del Nord mostrano un'elevata capacità di trattenere i propri laureati, con tassi di permanenza sempre largamente superiori al 70%, nel Sud e nelle Isole la dinamica è differente. Nello specifico, tra i laureati magistrali del Sud, oltre un

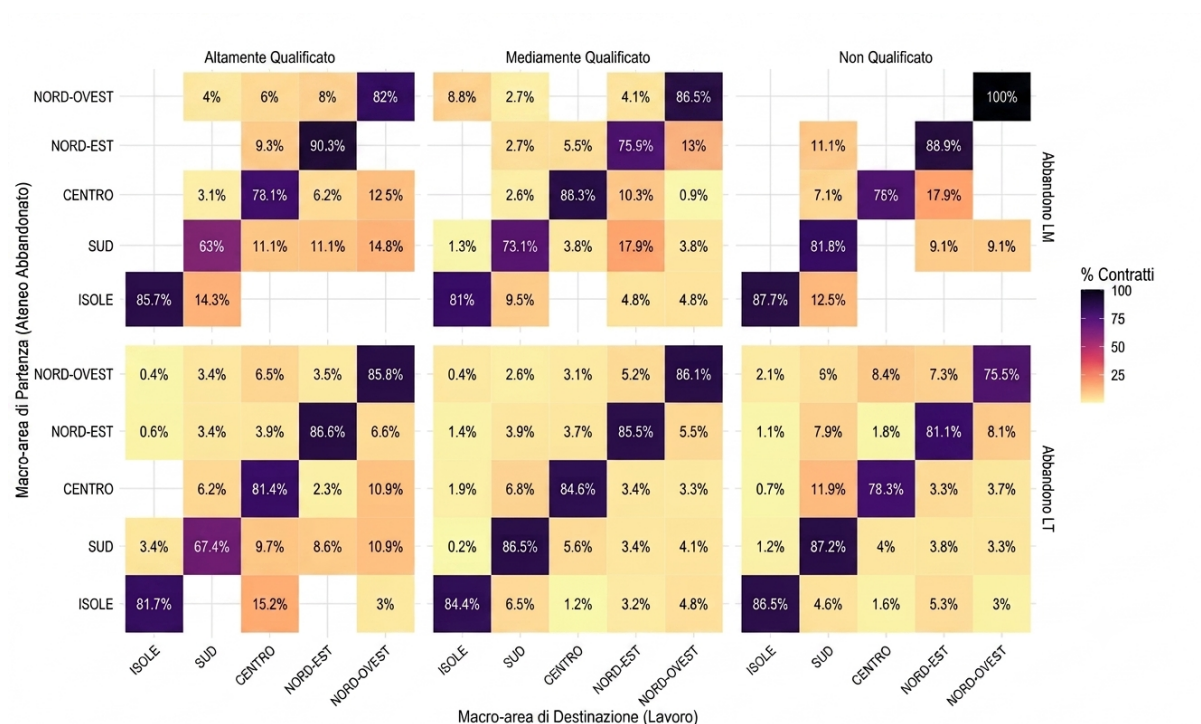
contratto su due tra quelli ad alta qualifica (il 53,2%) viene stipulato in un'area diversa da quella di studio, con una marcata polarizzazione verso il Nord-Ovest.

In generale, il Centro-Nord e in particolare il Nord-Ovest si configurano come i principali poli di attrazione, anche per le professioni non qualificate. È il caso, ad esempio, dei laureati magistrali delle Isole: dell'8,9% che svolge un lavoro non qualificato (cfr. tabella A3 in Appendice), il 60% lavora in regioni del Nord-Ovest. Invece, i tassi di permanenza nella macroarea di origine relativi al lavoro mediamente qualificato superano ovunque il 75%, suggerendo una buona capacità di assorbimento di questo tipo di professioni da parte dei mercati locali.

Di particolare interesse è il fenomeno della mobilità di ritorno, con cui si identificano gli individui che – al termine della loro esperienza universitaria fuori sede – tornano a lavorare nella area di origine. Nel nostro campione, il 41,4% dei laureati triennali che avevano effettuato uno spostamento di area rientra nella macro-area di provenienza in almeno

7 Per le analisi sul lavoro dei laureati sono stati considerati i soli contratti stipulati almeno nello stesso anno di ottenimento della laurea o anni successivi, in modo da valutare l'effetto del conseguimento del titolo.

Figura 3. Quote percentuali dei rapporti di lavoro per macro-area di partenza (sede dell'ateneo) e di arrivo (luogo di lavoro) e livello di qualificazione del lavoro (classificazione ISCO), per gli studenti che hanno abbandonato il percorso universitario



Nota: Panel superiore: laurea magistrale, panel inferiore: laurea di primo livello. Coorti di immatricolazione 2009/10-2016/17. Rapporti di lavoro osservati negli archivi WHIP/COB fino al 2019.

Fonte: elaborazione degli Autori

una fase della carriera lavorativa. Tra gli studenti originari del Sud e delle Isole – che sperimentano la mobilità di studio di gran lunga più elevata – il tasso di ritorno è molto più basso rispetto a quello osservato nelle aree del Centro e del Nord: pari rispettivamente al 43,4% e 47,3% per i laureati triennali e al 26% e 29,3% per i laureati magistrali. Nelle aree del Centro e del Nord, invece, i tassi di ritorno si attestano intorno al 60-80% per i laureati triennali e tra il 43% e il 72% per i laureati magistrali (tabella A4 in Appendice).

Il grafico relativo agli abbandoni (figura 3) mostra un pattern strutturalmente diverso: chi abbandona il percorso universitario tende a restare nella macro-area dell'ateneo in misura maggiore rispetto a chi lo conclude. Per gli abbandoni di percorsi di laurea magistrale, i tassi di permanenza superano quasi ovunque l'80-85%, evidenziando una ridotta propensione alla mobilità tra coloro che non conseguono il titolo. I casi di mobilità risultano, anche in questo caso, polarizzati verso il Nord: in particolare, si osservano spostamenti dal Sud verso il Centro e il Nord-Ovest per lavori altamente qualificati e

dal Sud verso il Nord-Est per quelli mediamente qualificati.

Nel caso degli abbandoni triennali, i tassi di permanenza sono più elevati e omogenei tra livelli di qualificazione. Fa eccezione, anche in questo caso, il Sud, dove la capacità di trattenere profili qualificati è sensibilmente minore. Va inoltre notato che i flussi verso professioni ad alta qualifica rappresentano una componente minoritaria del fenomeno: nel caso dei drop-out triennali, meno del 5% degli individui che hanno studiato al Sud riesce a inserirsi in segmenti lavorativi qualificati. Diversamente, al Nord oltre al 10% dei contratti è stipulato per lavori altamente qualificati, anche in assenza di una laurea (cfr. tabella A3 in Appendice).

In sintesi, il completamento del percorso universitario sembra modificare la natura della mobilità territoriale: nei laureati essa assume carattere selettivo e orientato verso segmenti qualificati del mercato del lavoro; tra coloro che hanno abbandonato il percorso di studi tende invece a configurarsi come mobilità meno strutturata e meno associata a occupazioni di elevata qualità, specialmente per chi non ha conseguito la laurea

triennale. L'interruzione degli studi, dunque, agisce da freno alla mobilità geografica.

Conclusioni

La sperimentazione presentata in questo contributo mostra come dati amministrativi prodotti da enti diversi possano essere integrati, anche in assenza di identificativi diretti, attraverso una procedura di *linkage* costruita nel rispetto della tutela della riservatezza. Attraverso un disegno a doppio cieco e l'uso di un set minimo di informazioni (sesso, data e comune di nascita), combinati con regole di esclusione temporali e biografiche basate sulla coerenza tra carriere accademiche e lavorative, è stato costruito un dataset di circa 17.000 match 1:1 ad alta affidabilità tra percorsi di studio e rapporti di lavoro. Questo sottoinsieme costituisce una base informativa utile per esplorare diversi temi di interesse, tra i quali la mobilità interna e il confronto tra percorsi di laureati e non laureati, con una profondità temporale e territoriale difficilmente raggiungibile con le rilevazioni attualmente disponibili. Al netto della forte dipendenza dalla dimensione del comune di nascita, l'appartenenza al gruppo 1:1 non è selettiva rispetto alle principali caratteristiche individuali e dei percorsi universitari considerati.

La struttura integrata del dataset consente infatti di sviluppare una lettura più granulare e dinamica delle traiettorie studio-lavoro, valorizzando appieno la dimensione longitudinale delle informazioni disponibili. In particolare, l'archivio ANS permette di seguire l'intero percorso universitario degli individui, dall'immatricolazione al conseguimento del titolo o all'abbandono, osservando esami sostenuti, crediti acquisiti e tempistiche di avanzamento. L'integrazione con i dati sulle carriere lavorative consente di collegare tali traiettorie formative agli esiti occupazionali successivi, aprendo la possibilità di analizzare non solo la qualità dell'inserimento lavorativo, ma anche la relazione tra progressione

accademica, interruzione degli studi e dinamica delle carriere nel medio periodo.

In questa prospettiva, il lavoro presentato apre alcune linee di sviluppo, principalmente orientate a migliorare la qualità del *linkage* e l'utilizzabilità del dataset risultante.

Un primo ambito riguarda la validazione del dataset dei match 1:1. Il confronto con *linkage* deterministici realizzati su base locale, laddove disponibili, consentirebbe di valutare in modo più diretto la qualità dei match 1:1 e di definire criteri più precisi per la loro inclusione nelle analisi empiriche.

Un secondo ambito riguarda l'estensione dell'analisi oltre il sottoinsieme dei match univoci. Il dataset 1:1 rappresenta infatti solo una parte dell'informazione disponibile. Una direzione promettente consiste nello sviluppo di strategie di assegnazione probabilistica e di imputazione multipla, basate su informazioni ausiliarie provenienti da altre fonti statistiche, per utilizzare anche i match non univoci (1:molti, molti:1 e molti:molti), incorporando così esplicitamente l'incertezza nelle analisi empiriche.

Infine, un aspetto cruciale riguarda la costruzione di infrastrutture dati stabili che rendano possibile, su base continuativa, l'integrazione tra archivi di istruzione e lavoro a fini di ricerca. La sperimentazione qui presentata dimostra la fattibilità tecnica di questa integrazione in un contesto privacy-preserving, ma evidenzia al tempo stesso la necessità di sviluppare soluzioni istituzionali e organizzative che ne consentano l'utilizzo sistematico su scala nazionale.

Nel complesso, il lavoro suggerisce che il valore analitico dei dati amministrativi non risiede solo nella loro disponibilità, ma nella possibilità di integrarli in modo metodologicamente solido, esplicito nei limiti e trasparente nelle assunzioni. In questa prospettiva, il *linkage* tra ANS e WHIP rappresenta non solo uno strumento per l'analisi empirica, ma anche un banco di prova per lo sviluppo di nuove infrastrutture di ricerca basate su dati amministrativi.

Appendice

Tabella A1. Tasso di risposta nell'indagine AlmaLaurea 2022 ai laureati

	N. laureati	% di risposta sul totale dei laureati	% di risposta sul totale dei laureati contattabili
A un anno			
Primo livello	167.467	68,7	72,6
Magistrali biennali	94.673	67,1	73,9
Magistrali a ciclo unico	34.257	69,2	73,9
A tre anni			
Primo livello	72.945	15,2	15,2
Magistrali biennali	86.117	53,7	76,8
Magistrali a ciclo unico	35.969	68,9	75,8
A cinque anni			
Primo livello	61.490	10,2	10,2
Magistrali biennali	80.369	55,3	64,5
Magistrali a ciclo unico	36.082	54,1	63,6
Sc. Formazione (pre-riforma)	409	54,8	60,4

Fonte: AlmaLaurea, XXV Indagine Condizione occupazionale dei laureati

Tabella A2. Probabilità di appartenenza al gruppo 1:1 rispetto a match non univoci. Risultati da modello logit

Variabile	(1)			(2)		
	AME	SE	p-value	AME	SE	p-value
Genere maschile	0,0291	0,0035	0,0000	0,0284	0,0034	0,0000
Coorte imm. (rif 2009)						
2010	-0,0045	0,0061	0,4569	-0,0046	0,0061	0,4522
2012	0,0024	0,0060	0,6903	0,0025	0,0061	0,6846
2013	-0,0257	0,0060	0,0000	-0,0250	0,0061	0,0000
2014	-0,0242	0,0061	0,0001	-0,0224	0,0064	0,0005
2015	-0,0298	0,0061	0,0000	-0,0275	0,0064	0,0000
2016	-0,0159	0,0064	0,0132	-0,0137	0,0069	0,0453
Tipo diploma (rif. Altri licei)						
Classico	-0,0106	0,0063	0,0920	-0,0114	0,0063	0,0676
Scientifico	-0,0113	0,0049	0,0217	-0,0119	0,0049	0,0141
Professionale	0,0070	0,0077	0,3634	0,0071	0,0077	0,3597
Tecnico	-0,0023	0,0053	0,6612	-0,0030	0,0052	0,5690
Esteri/Altro	-0,0246	0,0141	0,0803	-0,0256	0,0141	0,0682
Dim. Comune (rif.<10 mila)						
10-25 mila	-0,0875	0,0109	0,0000	-0,0873	0,0109	0,0000
25-50 mila	-0,1883	0,0103	0,0000	-0,1885	0,0103	0,0000
50-100 mila	-0,3416	0,0103	0,0000	-0,3418	0,0103	0,0000
100-250 mila	-0,5093	0,0101	0,0000	-0,5095	0,0101	0,0000
250-500 mila	-0,6515	0,0098	0,0000	-0,6517	0,0098	0,0000
500-3 milioni	-0,7164	0,0090	0,0000	-0,7168	0,0090	0,0000

segue

segue tabella A2

Area disciplinare (rif. Sanitaria)			
Economico-giuridica	0,0071	0,0074	0,3337
Formazione	0,0321	0,0115	0,0051
Ingegneria-Architettura	0,0099	0,0076	0,1921
Politico-sociale	0,0067	0,0090	0,4562
Scienze-tecnologia	0,0271	0,0078	0,0005
Umanistico-linguistica	0,0156	0,0075	0,0384
Non laureati	0,0107	0,0059	0,0729
Stato percorso (rif. Laurea Triennale in corso)			
Laurea Triennale	-0,0031	0,0052	0,5433
Laurea Magistrale	0,0093	0,0060	0,1210
Laurea Magistrale in corso	0,0051	0,0057	0,3728
Abbandono Laurea Triennale	0,0035	0,0050	0,4827
Abbandono Laurea Magistrale	0,0113	0,0143	0,4307
Passaggio a Laurea a Ciclo Unico	-0,0232	0,0187	0,2145
Fonte: elaborazione degli Autori			

Tabella A3. Qualifica dell'occupazione per area geografica dell'Ateneo e percorso universitario

	Altamente qualificato	Mediamente qualificato	Non qualificato
Laureati magistrali			
Nord-Ovest	63,2	30,9	5,9
Nord-Est	56,3	34,5	9,2
Centro	56,6	34,6	8,8
Sud	55,1	35,5	9,4
Isole	56,3	34,8	8,9
Laureati triennali			
Nord-Ovest	36,4	50,2	13,4
Nord-Est	26,3	57,0	16,7
Centro	26,5	61,0	12,5
Sud	22,4	59,0	18,6
Isole	23,6	55,1	21,3
Abbandoni magistrali			
Nord-Ovest	38,2	56,5	5,3
Nord-Est	33,0	57,4	9,6
Centro	18,1	66,2	15,7
Sud	23,3	67,3	9,4
Isole	19,5	58,4	22,1
Abbandoni triennali			
Nord-Ovest	12,3	69,9	17,8
Nord-Est	9,3	67,6	23,1
Centro	6,5	72,9	20,6
Sud	5,5	68,2	26,3
Isole	3,5	64,0	32,5

Fonte: elaborazione degli Autori

Tabella A4. Quota di studenti fuori sede e tasso di ritorno al termine degli studi per area di origine

Area di origine	% Fuori sede	Tasso di ritorno
Laureati triennali		
Nord-Ovest	13,8	79,4
Nord-Est	9,8	58,3
Centro	13,8	67,6
Sud	40,2	43,4
Isole	22,4	47,3
Laureati magistrali		
Nord-Ovest	8,0	72,5
Nord-Est	8,7	47,7
Centro	16,1	43,2
Sud	48,9	26,0
Isole	18,3	29,3
Abbandoni triennali		
Nord-Ovest	11,0	63,9
Nord-Est	6,4	90,5
Centro	11,3	67,6
Sud	54,6	61,8
Isole	16,8	62,0
Abbandoni magistrali		
Nord-Ovest	9,1	33,3
Nord-Est	30,3	70,0
Centro	9,1	66,7
Sud	39,4	25,0
Isole	12,1	53,8

Fonte: elaborazione degli Autori

Bibliografia

- Ardito C., Cottini E., Moroni G., Serrano-Alarcon M. (2026), *Ageing on the Job: How Raising Retirement Age Affects Work Accidents*, WorkInps paper n.110, gennaio
- Attanasio M., Ragozini G., Porcu, M., Giambalvo O. (2020), *Verso nord: Le nuove e vecchie rotte delle migrazioni universitarie*, Milano, Franco Angeli
- Berton F., Richiardi M., Sacchi, S. (2012), *The political economy of work security and flexibility: Italy in comparative perspective*. Policy Press, Bristol
- D'Agostino A., Ghellini G., Longobardi S. (2019), Out-migration of university enrolment: the mobility behaviour of Italian students, *International Journal of Manpower*, 40, n.1, pp.56-72
- De Angelis I., Mariani V., Torrini R. (2017), New evidence on interregional mobility of students in tertiary education: The case of Italy, *Politica economica*, 33, n.1, pp.73-96
- Devicienti F., Grinza E., Manello A., Vannoni D. (2025), Employer cooperation, productivity and wages: new evidence from inter-firm formal network agreements, *Economica*, 9, n.365, 1-41
- Dotti N. F., Fratesi U., Lenzi C., Percoco M. (2013), Local labour markets and the interregional mobility of Italian university students, *Spatial Economic Analysis*, 8, n.4, pp.443-468

- Galizzi M., Leombruni R., Pacelli L. (2023), Severe work disabilities and long-lasting losses, *Labour Economics*, 85, Article n.102432, <<https://doi.org/10.1016/j.labeco.2023.102432>>
- Leombruni R. (2024), Interviewing administrative records. A conceptual map for the use of big data for economic research, *Italian Journal of Applied Statistics*, 36, n.3, pp.295-326
- Leombruni R., Ardito C. (2026), The long and winding road. The access to administrative data for research in Italy, in Berton F., Paruolo P. (a cura di), *Data-Driven Learning in the EU. From Administrative Sources to Public Value*, Springer Cham, Joint Research Centre
- Priulla A., Miaci E., Panichella N. (2025), Social selection in student mobility: The interplay of origin, achievement and track in the Italian North-South divide, *Research in Social Stratification and Mobility*, 100, n.101105, pp.1-10
- Tosi F., Impicciatore R., Rettaroli R. (2025), Italian graduates' geographical mobility patterns: selectivity and regional inequalities, *Higher Education*, n.91, pp.869-893
- Usala C., Porcu M., Sulis I. (2023), The high school effect on students' mobility choices, *Statistical Methods & Applications*, 32, n.4, pp.1259-1293

Andrea Priulla

andrea.priulla@unikore.it

Ricercatore in Statistica presso il Dipartimento di Scienze economiche e giuridiche dell'Università degli Studi di Enna "Kore". La sua attività di ricerca si concentra sulle disuguaglianze socioeconomiche, di genere, etniche e territoriali nell'istruzione universitaria italiana. Tra le pubblicazioni recenti si segnalano: con Albano A., D'Angelo N., Attanasio M., Predicting university enrollment choices in Italy: a machine learning analysis of high school background and gender differences, *Higher Education*, 2025 e con Miaci E., Panichella N., Social Selection in Student Mobility, The Interplay of Origin, Achievement and Track in the Italian North-South Divide, *Research in Social Stratification and Mobility*, 2025.

Dalit Contini

dalit.contini@unito.it

Professoressa ordinaria di Statistica sociale presso il Dipartimento di Economia e statistica "Cognetti de Martiis" dell'Università di Torino. Si occupa di disuguaglianze educative, istruzione terziaria, valutazione di politiche educative. Tra le pubblicazioni recenti si segnalano: con Di Tommaso M.L., Muratori C., Piazzalunga D., Schiavon L., A lost generation? Impact of Covid-19 on high school students' achievements, *Journal of Economic Inequality*, 2025 e con Salza G., Children left behind. New evidence on the (adverse) impact of grade retention on educational careers, *European Sociological Review*, 2024.

Sonia Della Monica

sonia.dellamonica@unito.it

Assegnista di ricerca presso il Dipartimento di Economia e statistica "Cognetti de Martiis" dell'Università di Torino. Fra le pubblicazioni recenti si segnalano: con Leombruni R., Infortuni lavorando, ma senza tutele. Nuove evidenze su sottocopertura e sotto-notifica degli infortuni sul lavoro in Italia, *Economia & lavoro*, 2026 e con Storti L., Going up: Romanian Immigrants in an Alpine Region, *Mondi Migranti*, 2026.

Noemi Ferrara

noemi.ferrara@edu.unito.it

Borsista di ricerca presso il Dipartimento di Economia e statistica "Cognetti de Martiis" dell'Università di Torino e ricercatrice junior presso l'Istituto di Ricerche economiche e sociali del Piemonte (IRES).